

<http://dx.doi.org/10.7236/IIBC.2016.16.1.273>

IIBC 2016-1-37

인터넷 매체 언어의 국어 파괴 현상의 고찰을 통한 표준어 자동 번역 기술에 대한 연구

Research on Automatic Translation of Standard Language through Contemplation of Korean Destruction Phenomena in Internet Media Language

박장혁*, 정재훈**, 김신령***, 김영곤****

Jang-Hyuk Park*, Jae-Hun Jung**, Sin-Ryeong Kim***, Young-Gon Kim****

요약 본 논문에서는 인터넷 매체 언어상에서 나타나고 있는 국어 파괴 현상의 고찰을 고찰하여, 이에 대한 기술적 개선 방법으로서 표준어 자동 번역 기술을 제안한다. 실생활 속에서 국어 파괴 현상과 관련된 경험 및 피해 사례들을 설문 조사 및 FGI(Focussed Group Interview)를 통하여 수집 및 분석한 결과를 제시하고, 분석결과 자료 기반으로, '국어 파괴도(degree of Korean destruction)'의 지표를 정의하였다. 국어 파괴도는 표준어 자동 번역 기술이 갖고 있는 번역 오류가 가진 한계를 최소화하여, 표준어 자동 번역 기술의 효율성을 최대로 활용하기 위한 척도로서 활용될 수 있음을 인터넷 매체 문장 2,480 개의 분석을 통하여 제시하였다. 본 논문에서 제안된 국어 파괴도 측정 및 표준어 자동 번역 기술은 Java 언어를 사용하여 REST API 형태로 구현하였으며, 웹브라우저상에서 동작을 확인하였다.

Abstract In this paper, we consider the discussion of the breakdown phenomenon of Korean language that is displayed in the language of the Internet media, as a technical improvement approach to this, and to provide an automatic translation technology of standard language. The collected through real life surveys and experience and damage case regarding language breakdown phenomena in the FGI (Focussed Group Interview), and presents the results of the analysis, based on the analysis result data, we defined the index of language destroyed (degree of Korean destruction). The limit of the national language break or automatic translation technology standard language marked with translation errors that have to be minimized, the efficiency of the automatic translation technology standard language which may be utilized as a measure to get the maximum, and presented through the Internet media texts 2,480 pieces of analysis. Automatic translation technology of the standard language and was also measured the proposed language destroyed in this paper, we implemented in the form of a REST API using the Java language, was confirmed to work on the Web browser.

Key Words : degree of Korean destruction, FGI, REST API, automatic translation technology, Internet media language

*준회원, 한국산업기술대학교 컴퓨터공학과

**준회원, 한국산업기술대학교 컴퓨터공학과

***정회원, 동서대학교 정보통신과

****정회원, 한국산업기술대학교 컴퓨터공학과

접수일자: 2015년 10월 22일, 수정완료: 2015년 12월 26일

게재확정일자: 2016년 2월 5일

Received: 22 October, 2015 / Revised: 26 December, 2015 /

Accepted: 5 February, 2016

****Corresponding Author: ykkim@kpu.ac.kr

Dept. of Computer Engineering, Korea Polytechnic University, Korea

I. 서론

최근 스마트 기기의 보급이 대중화되면서 생활속의 인터넷 접속 시간과 사용 빈도수가 점점 늘어나고 있다. 이와 같은 인터넷 접속 문화의 확산은 신분의 노출이 제한적이고, 입출력 장치의 한계, 커뮤니케이션 제어 기능의 부재라는 인터넷 문화의 특성에 기반하여 인터넷 상에서의 언어 파괴 현상을 낳게 되었다. 인터넷 문화에 기반하여 사용되어지고 있는 언어를 ‘인터넷 매체 언어’라고 정의하며, 인터넷 매체 언어상에서 발견되어 지고 있는 언어의 본질적 문법 규범의 파괴 현상은 ‘국어 파괴 현상’이라고 정의된다^[1]. 국어 파괴 현상은 인터넷의 아버지라고 불리는 Tim Berners-Lee가 제창한 처음 월드 와이드 웹을 제창할 당시의 본연의 목표인 ‘소통과 공유’를 어렵게 하는 근본적 원인으로 지적되고 있다^[2]. 본 논문에서는 이러한 인터넷 매체 언어상에서 발생하고 있는 국어 언어 파괴 현상을 고찰하여, 표준어로 자동 번역하는 기술적 방법을 제안하고자 한다.

II. 이론적 배경

1. 국어 파괴 현상과 관련된 인터넷 매체, 언론 보도

가. “암호에 가까운 ‘10대들의 언어파괴’ 심각” [3]

국립국어원이 최근 ‘청소년 언어실태, 언어의식 조사’를 한 결과 초·중·고 재학생 95%가 일상생활 대화에서 신조어와 욕설을 섞어 쓰고 있는 것으로 나타났다. 최근 심각한 사회 현상인 학교폭력의 대부분이 사실상 언어폭력으로부터 시작되며, 아이들이 심한 욕설을 주고받는 모습을 보고 “욕설을 들었을 때 적극 훈계했다”는 교사는 43.9%에 그쳐 이에 대한 대책 마련이 시급한 실정이다. 중앙일보 어문연구소 부국장은 “10대들의 언어는 속도와 축약을 특징으로 하는 문자메시지의 영향으로 더욱 확대되는 경향이 있으며 그들만의 결속을 다지는 역할을 하기도 한다”면서 “그러나 다양한 어휘를 사용하지 못함으로써 어휘력의 빈곤을 드러내고 나아가 사고력의 부족을 초래하기도 한다”고 말했다.

나. “노잼, 열폭, 낫닝겐? . . . ‘한글 파괴’ 어디까지?” [4]

대통령 소속 국민대통합위원회가 지난해 1월부터 올해 4월까지 청소년들이 인터넷 공간에 올린 게시물 13만 건을 빅 데이터로 분석했다. 이 가운데 32%가 욕설이나 줄임말 같은 은어로 채워져 있었다. 조사 대상자 절반 이상이 은어의 부작용을 걱정했고, 세 명 중 한명은 세대 간 ‘언어 장벽’을 느낀다고 답했다.

다. “‘음오아에 ‘.’ 심쿵해 ‘.’ 째어 ‘. . . 국어파괴 vs 또다른 소통’ [5]

‘음오아에, 심쿵해, 째어...’ 뜻을 알 듯 모를 듯한 이 단어들은 요즘 인기 있는 대중가요의 제목이다. 인터넷 채팅방, 모바일 메신저 등에서 주로 볼 수 있었던 단어가 대중의 공감을 이끌어내는 대중음악에까지 흘러들어온 것을 두고 ‘국어 파괴’라는 지적이 나온다. 그러면서도 일부에서는 신조어가 당대 사회를 비추는 거울의 구실을 하는 만큼 ‘새로운 문화로서 무시할 수만은 없다’는 의견도 제기하고 있다. ‘어문법적으로도 맞지 않는 것은 물론 언어의 미학도 떨어지는 단점도 있다’고 지적했다.

라. 보도 내용의 고찰을 통한 정리/비교 결과

표 1. 보도 내용 정리 및 비교 결과

Table 1. Arrangement and comparison of information reported

번호	지적하는 문제	개선 방안	기사의 특징
(1)	언어 폭력	교육청 주관 예방	효과에 대해 비판적
(2)	비속어 / 소통 단절	-	
(3)	언어의 미학이 떨어짐	-	대중문화에서도 신조어를 사용하고 있는 실태를 보도함

각 보도 내용들을 정리 및 비교한 결과가 표 1에 기술되어 있으며, 검색을 통하여 확인되는 보도의 빈도수가 2015년 들어 급격히 증가하고 있음을 확인할 수 있다. 국어 파괴 현상에 대한 부정적 기능으로서 세대간 소통의 단절과 우리 문화 전반에 미치는 부정적 영향을 많이 언급하고 있으며, 긍정적 기능으로서 ‘감정 해소의 도구’라는 시각도 확인할 수 있었다.

2. 국어 파괴 현상과 관련된 기존 연구활동

가. 학술 논문/학위 논문 동향

- (1) ‘익명성’이 인터넷 게시물 제목에서의 ‘욕설/비속어’ 사용과 어떤 관계에 있는지를 통계학적으로 고찰하여 분석한 연구 내용을 발표하였다^[6]. 49개 자유 게시판의 글 제목 6,000 여개를 대상으로 ‘실명제’ 여부와 욕설/비속어 사용 비율을 분석하고, ‘실명제 게시판’과 ‘비실명제 게시판’의 욕설/비속어 표현 정도를 비교함으로써 익명성과 거친 표현의 상관관계를 살펴보고, 이러한 조건에 따른 통신 언어 쓰임의 변이를 구체적으로 밝혀내었다.
- (2) 한국어와 영어로 이루어지는 인터넷 통신 의사소통에서 나타나는 남성과 여성의 성차와 성역학 관계에 대하여 논의하는 연구를 수행하였는데, 각 성차에 의한 의사 표현 시에 나타나는 담화 분석에서 인터넷 매체 언어가 어떠한 국어 파괴 현상의 영향을 받고 있는지에 대한 다양한 근거를 확인할 수 있다^[7]. 예를 들어, 여성들의 경우 ‘겸양의 1인칭 대명사’인 ‘저’를 빈번하게 사용하는 한편, 남성의 경우, ‘대명사’인 ‘나’가 생략되는 문법 변형 문장이 많음을 주장하였다.
- (3) 대구 지역 대학생들의 인터넷 게시글을 대상으로 언어 사용 실태를 파악하고, 인터넷 매체 언어의 영향이 얼마나, 어떻게 나타나고 있는지를 살펴본 연구를 수행하였다^[8]. 해당 연구를 통하여 전체 분석글의 83%에서 국어 파괴 현상이 나타나고 있음을 확인하였으며, ‘대자보’와 같은 대학생들의 실제 게시물에서 이처럼 높은 빈도로 인터넷 매체 언어가 일상적으로 사용되고 있음을 주장하였다.

나. 관련 저서 동향

- (1) 네티즌의 문화와 언어를 새로운 문화 현상으로 조망하고, 네티즌의 문화와 언어를 다양한 측면에서 다룬 개론적인 성격의 저서를 발표하였다^[9]. 저자는 해당 저서에서 인터넷 매체 언어를 소통 양식, 특징, 유형, 어휘, 문법, 담화(대화), 해외 네티즌 언어의 분류로 분석하여 다양한 사례를 들어 논의하였다.
- (2) 정통 문법을 공부하고 언어학을 전공한 학자의 입장에서 바라본 문자메시지와 통신 언어에 대한 전반적인 현상을 논의한 저서를 발표하였다^[10]. 영어 문장에서 나타나고 있는 이모티콘 사용 현상에 대한 언어학자의 시각을 논의하고 있다.

III. 표준 번역 프로세스

1. 국어 파괴 현상의 분류 통합을 위한 근거

본 논문에서는 이정복^[2, 8]에서 분류한 국어 파괴 현상의 분류를 기준으로, 최니아^[11]의 연구에서 언급된 분류 항목 중 ‘신조어’, ‘비속어’, ‘의성어’, ‘외계어’를 추가하여 국어 파괴 현상의 항목을 기준으로 인터넷 매체 문장의 국어 파괴 현상을 정리하였다. 항목 정리 결과는 분류 항목 기준 중, 중복되는 항목과 상호 포함 관계에 있다고 판단되는 항목을 정리하였고, 항목 정리를 위한 판단 근거는 표 2와 같다.

표 2. 인터넷 매체 언어상에서 발생하고 있는 국어 파괴 현상의 분류

Table 2. Classification of korean destruction phenomena on the internet media language

국어 파괴 형태	의미	예문
기호의 사용	이모티콘, 숫자, 자모의 특정 패턴의 사용을 통한 의미 전달, 또는 의미 없는 습관적 사용.	~; TT 8282, 2000 츄크츄크
음운 형태 변이	음운을 첨가하거나 교체하여 표현함	추꾸를 사랑하시는 분(축구) / 알취(알지) / 따랑해(사랑해) / 나머지 아찌~~(알았지)
형태 통사	서술어 줄이기/문장 성분 줄이기 등의 형태 통사적 변화를 문장에 가미함	첨오시는 분들 마니마니환영(합니다)!! 다들 오세용! 여기롱!
신조어	사회 현상을 반영하여 새로 생긴 말	된장녀(문화적 사대주의와 개념없는 소비 행태를 보이는 여성)
비속어	통속적으로 쓰는 저속한 말	쌍까다(모른 척하고 무시) / 담탱이(담임교사)
의성어 / 의태어	소리나 모양을 나타내는 말로 감정을 표현	아햏햏(모호한 느낌) / 휘리릭(통신 상황에서 재빨리 사라지는 모습)
외계어	알파벳, 특수 문자 등을 혼합한 정제 불명의 표현	☆로 (별로) / ©랑웨(이렇게) / 말하지 않아도(않아도)

표 3. 항목 분류 통합을 위한 근거

Table 3. Basis for items integrated classification

좌나야	이정복	분류	판단 근거
이모티콘	그림 글자		이모티콘, 숫자는 자연어 처리 시에 '기호'로 동일 분류하여 처리할 수 있음
숫자/자모 표기	그림 글자	기호의 사용	
통신용 인사말	-	신조어	통신용 인사말(하이루)도 통신 세대 및 문화에서 기반한 신조어로 파악함
신조어	-		
비속어	-	비속어	-
줄임말	형태 통사	형태 통사	-
소리 나는 대로 쓰기	음운 형태 변이	음운 형태 변이	-
의미 변이	-	-	의미 변이와 품사 변이의 경우, 표준어 번역 단계를 거치지 않아도 의미 전달이 어느 정도 가능하다고 판단됨(도배하다/피가다 등)
품사 변이	-		
의성어/의태어	-		
외계어	-	외계어	-

2. 국어 파괴도의 정의 및 측정

본 논문에서는 국어 파괴 현상이 발생한 문장에서의 파괴 정도를 정의하여 자동 번역 알고리즘 설계시에 사용하고자 한다. 국어 파괴도를 정의하고자 하는 목적은 표준어 자동 번역 대상이 되는 문장을 식별하기 위해서이며 이러한 전처리 과정의 필요성은 일반적인 자동번역 기술이 가지고 있는 번역의 한계와 오용을 줄이고자 하는 것이다. 즉, 언어 파괴가 일정 수준 이하로 발생하고 있는 문장들의 경우, 자동 번역 대상 문장으로 간주하지 않음으로써 자동 번역의 오작동 범위를 줄이고자 한다.

국어 파괴 현상과 직접적인 관계가 있는 가독성 영향 지표로 로그우도(log likelihood: LL)를 주목하였다. 로그우도는 일반적인 글이 갖고 있는 어휘적인 요소로서 해당 글이 언어 현실속에서 자주 사용되는 단어들로 구성되어 있는지를 반영하는 지표이다.

국어 파괴 현상이 발생한 경우, 언어 현실과 괴리가 있는 어휘들이 사용되어 지므로 로그우도의 값이 영향을 받을 것이라는 직관 하에 국어 파괴도 지표를 정의하였다. 일반적인 로그우도는 다음과 같이 확률값을 기반으로 정의되어진다.

표 4 가독성 결정 요소와 가독성의 관계

Table 4. Readability determinants element and relationship between readability

스마트폰		태블릿	
요소	값	요소	값
글의 길이(TL)	-0.257	문장 당 평균 명사구 수(NP)	0.335
문장당 평균 대명사수(PRP)	0.104	문장 당 평균 단어수(WS)	0.276
로그우도(LL)	-0.100	문장 길이(SL)	0.234
어절당 어절 길이(CW)	-0.096	문장 당 평균 동사구 수(VP)	0.212

$$\prod_w P(w|M)^{C(w)} \tag{1}$$

$P(w|M)$ 는 주어진 말뭉치 M 에 대한 단어 w 가 나올 확률이고, $C(w)$ 는 단어 w 가 나타나는 횟수로 정의될 수 있다. 상기 수식에 대한 로그우도는 상기 식에 \log 식을 취하여 하기와 같이 정리된다.

$$\sum_w C(w) \log(P(w|M)) \tag{2}$$

본 논문에서는 말뭉치 내에 단어 w 가 존재하는 확률 및 횟수 대신에, 단어 w 의 형태소 분석 여부를 확인하여, 주어진 문장의 형태소 분석 단어 개수를 이용하여 상기 수식을 변형하여 사용하기로 한다. 근거는 실생활에서의 언어 현실의 자료 집합체인 말뭉치내의 단어 존재 확률은, 해당 단어의 형태소 분석 여부와 관련이 있기 때문이다.

본 논문에서는 로그우도와 같은 역할을 수행할 수 있는 지표로서 형태소 분석기의 분석 결과인 형태소 태깅 결과를 수식화하여 국어 파괴도를 정의하였고, 주어진 문장 S 의 국어파괴도 $D(S)$ 는 다음과 같이 정의된다.

$$D(S) = \left(\frac{\sum_u \alpha_u * S_u}{\sum_p S_p} \right) * 100 \tag{3}$$

상기 식에서, p 는 문장 S 에 속한 모든 형태소 태그들을 의미하며, $\sum_p S_p$ 는 총 형태소 분석 태그의 개수를 의미한다. 또한, u 는 형태소 분석이 되지 않은 태그(/unknown)들을 의미하며, α_u 는 '/unknown' 태그들의 문장 S 내의 위치에 따른 가중치를 의미한다. 이와 같이 정의된 의미적 분석에 대하여 표 5에 정리하여 기술하였다.

표 5. 국어 파괴도 지표 설명

Table 5. Indicators description of Korean destruction

국어 파괴도	의미	가중치의 의미
형태소 태그 정보에 기반하여 정의	<ul style="list-style-type: none"> 형태소 분석시, 분석 불가 태그(UN)의 개수를 위치에 따른 가중치를 주어 계산에 이용 형태소 분석 불가 태그(UN) 단어들의 경우, 문장 내에서의 위치에 따라, 분석 불가 태그(UN)가 문장 첫부분과 끝부분에 나타나면 3, 문장 중간에 나타나면 1의 가중치를 적용 	일반적으로 신조어, 비속어가 문장 끝부분에 나타날 경우, '동사구'의 역할을 하는 경우가 많으며, 이 경우에는 문장 전체 해석의 난이도를 높인다고 판단. 또한, 문장 첫부분에 나타날 경우, '주어부'가 되므로, 이 경우에도 문장 전체 해석의 난이도를 높인다고 판단

웹사이트에서 임의의 문장에 대한 국어파괴도 계산 결과를 확인할 수 있도록 개발을 진행하였으며, 개발 언어는 Java 프로그래밍 언어를 이용하여 구현하였다.

개발된 국어파괴도 웹사이트를 이용하여, 연구팀의 지인들로부터 수집한 카카오톡 2,480 문장에 대한 국어파괴도의 계산을 수행하였으며, 샘플 문장 2,480 개에 대한 국어파괴도의 분포 결과는 그림 1과 같다. 2,480 문장에 대한 평균 국어 파괴도는 35.005를 기록하여, 평균 3 단어가 이상이 형태소 분석이 이루어지지 않음을 발견하였고, 샘플 문장에 대한 범위별 국어 파괴도 분포는 그림 2에 도식화 하였다.

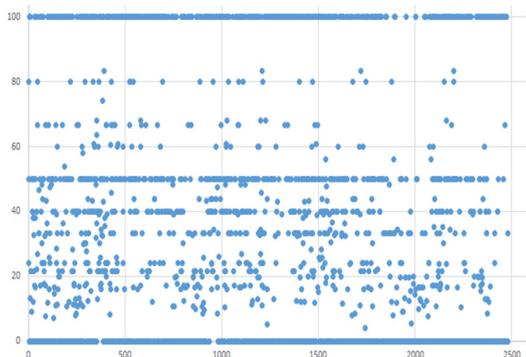


그림 1. 카카오톡 샘플 수집 문장 파괴도 분포
 Fig. 1. Destruction distribution with KakaoTalk sample collection sentences

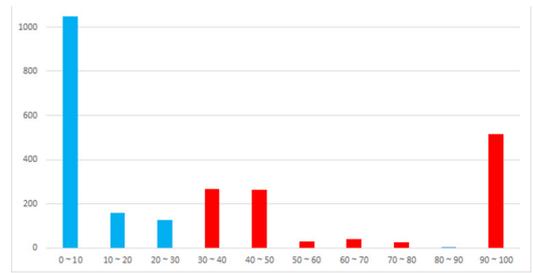


그림 2. 카카오톡 샘플 수집 문장 국어 파괴도의 범위별 분포
 Fig. 2. Ranges unit Distribution of Korean destruction with KakaoTalk Sample collection sentences

샘플 문장의 분석 결과, 30 미만의 국어 파괴도를 기록한 문장들의 경우, 대부분 '기호의 사용' 문장(예: "그건 살수있나ㅠ", "시간있을때 여행이나 다녀" 등)로 나타났으며, 전체 문장의 해석에 무리가 없는 경우가 많음을 발견하였다. 국어 파괴도 30 이상의 문장들을 표준어 자동 번역 대상 문장으로 설정하였다.

3. 자동 표준어 번역 알고리즘

국어 파괴도 30 이상의 문장들을 대상으로, 다음과 같은 단계를 거쳐 자동 표준어 번역을 수행하였으며, 자동 표준어 번역의 각 단계 설명을 돕기 위하여, 예제 문장으로 다음 문장을 사용하여 설명한다.

뜨아.. 여기 모이삼 저 영화 줌 홈쑼무인듯 ㅋㅋ

상기 문장의 국어 파괴도는 46.027 로서, 본 번역 알고리즘의 대상 문장이다. 상기 문장의 형태소 분석 결과는 그림 3과 같으며, 문장의 국어 파괴도 측정 화면은 그림 4와 같다.

```

morphem: "<![CDATA[@/SW{/SS이모/NNG타/NNG콘/NNG}/SS]]>"
}
-final_affect : {
  evidence: "ㅋㅋ"
  affect: "즐거움"
  valence:5.89
  arousal:5.54
}
morphem:"뜨아/UN../SW 여기 /NP 모이/VV삼/EFN 저/NNG 줌/MAG 홈쑼무인듯/UN @/SW{/SS이모 /NNG타/NNG콘/NNG} /SS"
input_str:"뜨아..여기 모이삼 저 영화 줌 홈쑼무인듯 ㅋㅋ"
    
```

그림 3. 예제 문장의 형태소 분석 결과
 Fig. 3. Morphological analysis results of the example sentence

Raw	JSON	Response
Copy to clipboard Save as file		
<pre>{ sentence: "쓰아.. 여기 모이삼 저 영화 좀 흠좀무인듯 ㅋㅋ" detractionRatio: 46.02700445097593 -parseResult: [8] 0: "쓰아/LN../SW" 1: "여기/MAG" 2: "모이/W삼/ERN" 3: "저/MMT" 4: "영화/WNG" 5: "좀/MAG" 6: "흠좀무인듯/LN" 7: "ㅋㅋ/LN" }</pre>		

그림 4. 예제 문장의 국어 파괴도 측정
Fig. 4. Measurement of Korean destruction with the example sentence

표준어 번역 알고리즘은 다음과 같은 4단계를 거쳐서 진행된다.

[1단계] 이모티콘, 숫자, 기호를 문장에서 분리한다. 분리된 이모티콘, 숫자, 기호는 국어 파괴 사전에 대응되는 문구가 있으면, 대응 문구로 변환되며 이모티콘의 경우 '감정'을 표현하는 단어이므로 해당 감성을 지문(地文)처럼 최종 문장에 삽입하는 방식을 취하였다. 상기 문장에서는 '/SE'로 태깅된 '...'이 분리하였다. '...'의 경우 습관적인 사용이 많은 기호로서 별도의 대응 단어 변환이 이루어지지 않으며, 'ㅋㅋ'의 경우, 국어 파괴 사전을 통하여 '즐거움'이란 감성을 인식하므로 최종 문장에 ('즐거움')이란 지문을 삽입하게 된다.

[2단계] '/UN'으로 태깅된 형태소의 경우, 국어 파괴 유형 중 형태 동사, 외계어, 신조어, 비속어일 가능성이 높으며 '국어 파괴 사전'을 대상으로 한 적절한 단어 검색을 시행한다. 상기 예문의 경우, '쓰아'와 '흠좀무'가 /UN으로 태깅되어 있으므로 각각 국어 파괴 사전을 검색하여, 감탄사의 경우, 해당 감성에 대한 지문으로(('놀람')), 신조어인 경우는 대응되는 표준어로 변환되는데, 표준어 변환시, 다음과 같은 경우를 고려하여 알맞은 변환을 시도한다.

- '/UN' 태그가 문장 첫 부분에 나타난 경우 : 이런 경우에는 '주어부'이거나 감탄사일 가능성이 높다고 판단하여 '감성 지문'이나 독립적인 절로 변환한다.
- '/UN' 태그가 문장 중간에 나타난 경우 : 대응되는 표준어로 변환한다.
- '/UN' 태그가 문장 끝부분에 나타난 경우 : '동사구'이거나 '감탄사'일 가능성이 높다고 판단하여 '감성

지문'이나 대응되는 동사절로 변환하고 상기 문장의 경우, '흠좀무'가 해당되며 '조금 무서운 것 같습니다'라는 동사절로 변환을 시도한다.

[3단계] 만약 '/UN' 태그 단어에 대하여 국어 파괴 사전에 등록되어 있지 않은 단어라면 '음운 형태 변이'를 검사한다. 국어 파괴 사전 검사 보다 '음운 형태 변이' 검사를 나중에 실시하는 이유는 '음운 형태 변이'의 경우, 10만 단어 이상으로 이루어진 형태소 분석 사전에서 검색을 수행해야 하는 관계로, [2단계] 검사를 통하여 검색 과정을 효율화할 수 있기 때문이다. '음운 형태 변이' 검사는 일반적인 문자열 검색 방법으로 접두사 또는 접미사를 대상으로 실시하며, Boyer-Moore 문자열 검색 방법 알고리즘을 사용하여 실시한다.

[4단계] 각 단계에서 변환에 사용되는 단어들을 해당 형태소 위치에 배치하여 문장을 재구성한다. 상기 예문의 경우, 다음과 같은 문장으로 변환된다.

(놀람) 여기 모이삼 저 영화 좀 조금 무서운 것 같습니다 (즐거움)

자동 표준어 번역 알고리즘을 도식화한 내용은 그림 5와 같다.

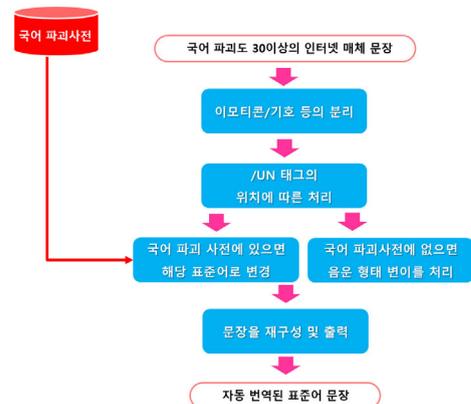


그림 5. 자동 표준어 번역 알고리즘
Fig. 5. Automatic translation algorithm with standard language

IV. 실험 및 결과

본 논문의 실험을 위한 알고리즘의 동작 방법은 7단계

를 거쳐 확인할 수 있다.

[1단계] 웹 브라우저를 실행한다. 본 논문에서는 Google사의 크롬(Chrome) 브라우저를 사용하였다.

[2단계] 크롬 브라우저의 메뉴 버튼을 클릭하여, “설정>확장프로그램”을 클릭한다.

[3단계] 확장 프로그램으로 REST API를 이용할 수 있는 클라이언트 플러그인을 다운받음. 본 논문에서는 “Advanced REST Client” 플러그인을 사용하였다.

[4단계] “설정>확장프로그램>Advanced REST Client” 항목의 “옵션”을 클릭한다.

[5단계] 옵션 클릭 후, 좌측 상단의 “Request”를 클릭하면, REST API의 다양한 명령을 실행할 수 있다.

[6단계] 상단 텍스트 박스에 REST API로 구현된 표준어 자동번역기의 URL을 입력한다.

[7단계] 하단의 “SEND” 버튼을 클릭하면 결과가 출력된다.

그림 6은 표준어 자동번역기가 예제 문장을 번역한 후 출력되는 화면이다.

```
input: "쓰아.. 여기 모이삼 저 영화 좀 홈쫌무인듯 ㅋㅋ"  
output: "쓰아..(놀람) 여기 모이삼 저 영화 좀 홈 이제 사실이라면 좀 무섭군요 ㅋㅋ(즐거움)"  
destructionRatio: {  
  sentence: "쓰아.. 여기 모이삼 저 영화 좀 홈쫌무인듯 ㅋㅋ"  
  destructionRatio: 46.02700445037593  
-parseResult: [s]  
  0: "쓰아/UN../SW"  
  1: "여기/MAG"  
  2: "모이/위삼/EFM"  
  3: "저/AMT"  
  4: "영화/MNG"  
  5: "좀/MAG"  
  6: "홈쫌무인듯/UN"  
  7: "ㅋㅋ/UN"  
}
```

그림 6. 표준어 자동 번역기의 문장 번역부
Fig. 6. Sentence translation unit of Automatic translator algorithm with standard language

준어 번역 기술을 제안하였다. 국어 파괴 현상의 고찰을 위하여 국내의 언론 보도 내용과 기존 연구 내용들을 고찰하였으며, 국어 파괴 현상에 대한 문제의식이 최근 들어 점점 더 증가되고 있는 경향을 확인할 수 있었다. 언론 보도 내용은 국어 파괴 현상에 대한 현재 상황을 보도 하면서 ‘문제의 심각성’을 알리는 내용이 주를 이루었으며, 이에 대한 사회적, 기술적 대응을 위한 논의에 대해서는 아직 부족한 모습을 확인하였으며, 기존 연구 역시 국어 파괴 현상에 대한 대응 방안에 대해서는 연구가 진행되고 있지 않다. 또한, 언론 보도 내용, 기존 연구 내용과는 별도로 실생활에 국어 파괴 현상이 미치고 있는 영향과 피해 사례들을 확인하였고, 설문 조사 결과를 통하여 국어 파괴 현상에 대한 피해 사례들과 이에 대한 기술적 대응의 필요성을 확인하였다. 기술적 대응 노력으로서 본 논문에서는 기계 번역 기술에 기반하여 ‘표준어 번역 기술’을 제안하였고, 기계 번역 기술의 한계인 번역의 정확도, 번역의 비용(cost)을 고려한 효과적인 활용을 위한 전처리 장치로서 ‘국어 파괴도(degree of Korean destruction)’이라는 국어 파괴 현상의 정량적 측정 지표를 제안하였다.

샘플 문장 2,480개를 통하여 ‘국어 파괴도’가 30이상 진행된 문장들에 대하여 ‘표준어 번역 기술’을 적용하는 것이 효과적임을 밝혔다. ‘표준어 번역 기술’은 형태소 번역이 이루어지지 않은 형태소들을 중심으로 국어 파괴 어휘 사전을 참조하여 번역하는 직접 번역 방식으로 알고리즘을 제안하였으며, REST API 형태로 구현하여 웹 브라우저를 통하여 검증하였다.

형태소 분석 정보에만 의존하고 있는 국어 파괴도 역시, 다양한 자연어 처리 정보들을 반영하는 형태로 확장되어질 수 있으며, 이러한 확장 방법에 대한 연구도 향후에 필요할 것으로 판단된다.

본 논문의 결과물을 메쉬업(MESH-UP) 형태의 웹서비스, 앱 서비스 개발을 지원할 수 있는 오픈 API 형태로 공개할 계획을 가지고 있으며, 이러한 노력들과 본 논문의 성과가 국어 파괴 현상에 대한 기술적 대응 노력에 대한 사회 및 학계의 관심을 불러일으키는데 기여할 수 있도록 계속 관심을 갖고 연구해나갈 것이다.

V. 결론

본 논문에서는 인터넷 매체 언어상에서 나타나는 국어 파괴 현상을 고찰하여, 기술적 대응 방법으로서의 표

References

[1] Na-ya Choi, Eugene Han, “Effects of Students’

Attitudes to Internet Media Language and Orthographic Knowledge on Their Use of Internet Media Language”, Korean Journal of Child Studies, 31(5), pp.31-45, 2010.

[2] Jeong-Bog Lee, “The expansion of the spread of Internet communication language and Korean studies”, Sotong, 2009.

[3] “Language destruction of ‘10 teenagers near the Password’ Serious“, yonhaptimes, 2014. 10. 8.

[4] “노잼 열풍 낫닝겐 ? ... Where to ‘Korean destruction’”, KBS news, 2015. 7. 14.

[5] “음오아에’·‘십쿵해’·‘쩨어’...Korean destruction vs Another communication“, sports dong-a, 2015. 6. 30.

[6] Semo-dol Son, “The Relation on Anonymity and Using swearwords/vulgar words in Titles on Net Boards”, The Text Linguistic Society of Korea, Text Linguistic, 15, pp.169-198, 2003.

[7] Kyoung-Sook Song, “An Analysis of Person Deixis in Korean and English Internet Chat”, The Text Linguistic Society of Korea, Text Linguistic, 15, pp.463-482, 2003.

[8] Jeong-Bog Lee, “The Internet Communication Language in Bulletin Board Messages by University Students of Daegu Area”, The Linguistic Society of Korea, Korean Linguistic, 21, pp.239-267, 2003.

[9] Eun-A Seo, Netizens Language, communication Books, 2007.

[10] David Crystal, Is it text messages the disaster of language? Is Evolution?, Almbbooks, 2011.

[11] Sung-Kwon Choi, Ki-Young Lee, Yoon-Hyung Roh, Oh-Woog Kwon, Young-Gil Kim, “Customization Method for Commercialization of a Pattern-based English-Korean Machine Translation System”, KOREA INFORMATION SCIENCE SOCIETY, Journal of KISS : Software and Applications, 39(4), pp.253-260, 2012.

[12] Sung-Kwon Choi, Young-Kil Kim, “Semantic Classification of Lexical Translation Patterns Extracted from Bilingual Corpus and Application

of Lexical Translation Patterns to MT System”, The Korean Association Of Translation Studies, The Journal of translation studies, 11(3), pp.277-301, 2010.

[13] M.G. Barry, M.E. Purcell, B.J. Eck, J. Hayes, E. Arandia, “Web Services for Water Systems: The iWIDGET REST API”, 16th Water Distribution System Analysis Conference, Procedia Engineering, Vol. 89, pp.1120-1127, 2014.

저자 소개

박 장 혁(준회원)



- 1992년 2월 : 숭실대학교 사회사업학과(학사)
- 1994.8년 : 숭실대학교 중소기업경영지도학과(경영학석사)
- 2011년 2월 ~ : 한국산업기술대학교 컴퓨터공학과 박사재학
- 1994년 ~ 2001년 : 중소기업연구원 책임연구원

<주관심분야 : 소프트웨어공학, 인터넷정보처리, 데이터 분석>

정 재 훈(준회원)



- 2010년 2월 : 한국산업기술대학교 컴퓨터공학과(공학사)
- 2012년 2월 : 한국산업기술대학교 IT 융합학과(공학석사)
- 2012년 3월 ~ : 한국산업기술대학교 컴퓨터공학과 박사재학

<주관심분야 : 소프트웨어공학, 정보통신시스템, 객체지향 분석 및 설계>

김 신 령(정회원)



- 1983년 2월 : 경북대학교 전자공학과(공학사)
- 1985년 2월 : 연세대학교 본대학원 전자공학과(공학석사)
- 1990년 2월 : 연세대학교 본대학원 전자공학과(공학박사)
- 1992년 2월 ~ : 동서대학교 정보통신과 부교수

<주관심분야 : 정보통신시스템, 부호화 방식>

김 영 곤(정회원)



- 1983년 2월 : 경북대학교 전자공학과 (공학사)
 - 1985년 2월 : 연세대학교 본대학원 전자공학과(공학석사)
 - 2000년 2월 : 한국과학기술원 전산학과(공학박사)
 - 1985년~2007년 : KT 수석연구원
 - 2007년 ~ : 한국산업기술대학교 컴퓨터공학과 부교수
- <주관심분야 : 소프트웨어공학, 정보통신시스템, 객체지향 분석 및 설계>